

UC Davis

UC Davis Previously Published Works

Title

Prevalent, Dynamic, and Conserved R-Loop Structures Associate with Specific Epigenomic Signatures in Mammals.

Permalink

<https://escholarship.org/uc/item/50v1z9wj>

Journal

Molecular cell, 63(1)

ISSN

1097-2765

Authors

Sanz, Lionel A
Hartono, Stella R
Lim, Yoong Wearn
et al.

Publication Date

2016-07-01

DOI

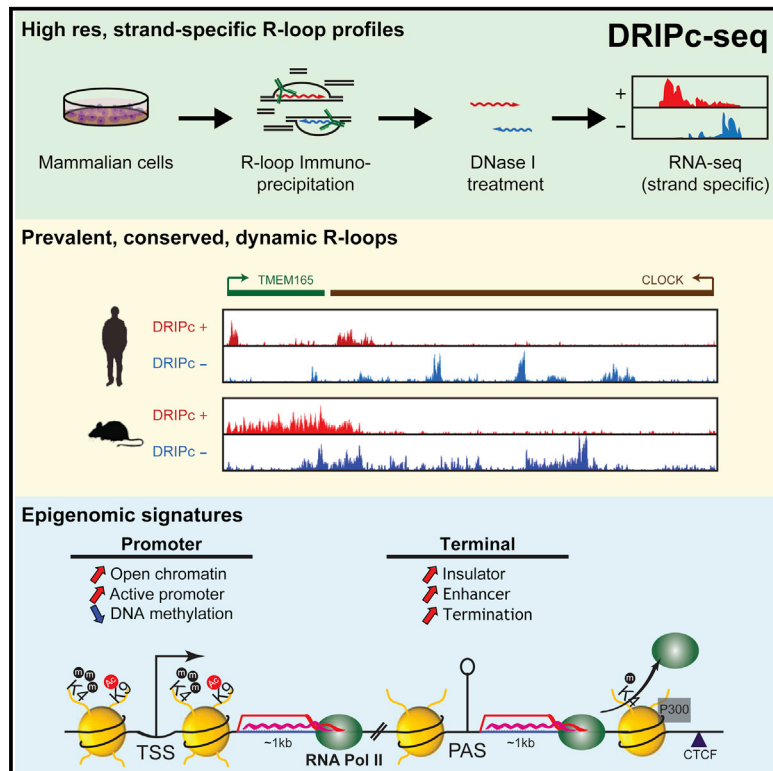
10.1016/j.molcel.2016.05.032

Peer reviewed

Molecular Cell

Prevalent, Dynamic, and Conserved R-Loop Structures Associate with Specific Epigenomic Signatures in Mammals

Graphical Abstract



Authors

Lionel A. Sanz, Stella R. Hartono, Yoong Wearn Lim, ..., Paul A. Ginno, Xiaoqin Xu, Frédéric Chédin

Correspondence

flchedin@ucdavis.edu

In Brief

Sanz et al. developed a high-resolution methodology to map R-loop structures and report that, in mammalian genomes, co-transcriptional R-loop formation is prevalent, conserved, and dynamic. The study highlights that promoters and terminators are hotspots of R-loop formation and associate with specific chromatin signatures.

Highlights

- Co-transcriptional R-loop formation is prevalent, conserved, and dynamic
- R-loops associate with specific epigenomic signatures at promoters and terminators
- Terminal R-loops associate with an insulator- and enhancer-like state
- Terminal R-loops define a new class of transcription terminators

Accession Numbers

GSE70189



Sanz et al., 2016, Molecular Cell 63, 167–178
 July 7, 2016 © 2016 Elsevier Inc.
<http://dx.doi.org/10.1016/j.molcel.2016.05.032>

CellPress

Prevalent, Dynamic, and Conserved R-Loop Structures Associate with Specific Epigenomic Signatures in Mammals

Lionel A. Sanz,^{1,3} Stella R. Hartono,^{1,3} Yoong Wearn Lim,^{1,3} Sandra Steyaert,² Aparna Rajpurkar,¹ Paul A. Ginno,^{1,4} Xiaoqin Xu,¹ and Frédéric Chédin^{1,*}

¹Department of Molecular and Cellular Biology and Genome Center, 1 Shields Avenue, University of California, Davis, Davis, CA 95616, USA

²Department of Mathematical Modeling, Statistics, and Bioinformatics, University of Ghent, Ghent 9000, Belgium

³Co-first author

⁴Present address: Friedrich Miescher Institute for Biomedical Research, Maulbeerstrasse 66, 4058 Basel, Switzerland

*Correspondence: fchedin@ucdavis.edu

<http://dx.doi.org/10.1016/j.molcel.2016.05.032>

SUMMARY

R-loops are three-stranded nucleic acid structures formed upon annealing of an RNA strand to one strand of duplex DNA. We profiled R-loops using a high-resolution, strand-specific methodology in human and mouse cell types. R-loops are prevalent, collectively occupying up to 5% of mammalian genomes. R-loop formation occurs over conserved genic hotspots such as promoter and terminator regions of poly(A)-dependent genes. In most cases, R-loops occur co-transcriptionally and undergo dynamic turnover. Detailed epigenomic profiling revealed that R-loops associate with specific chromatin signatures. At promoters, R-loops associate with a hyper-accessible state characteristic of unmethylated CpG island promoters. By contrast, terminal R-loops associate with an enhancer- and insulator-like state and define a broad class of transcription terminators. Together, this suggests that the retention of nascent RNA transcripts at their site of expression represents an abundant, dynamic, and programmed component of the mammalian chromatin that affects chromatin patterning and the control of gene expression.

INTRODUCTION

R-loops are three-stranded nucleic acid structures composed of an RNA:DNA hybrid and a displaced single-stranded DNA (ssDNA) loop. These non-B DNA structures have long been considered to result from rare and accidental entanglements of RNA with DNA during transcription. Aberrant R-loop formation has been linked to increased mutagenesis, hyper-recombination, rearrangements, and transcription-replication collisions as a result of topological, replication, or transcriptional stress (Aguilera and García-Muse, 2012; Hamperl and Cimprich, 2014). R-loop dysfunction is also thought to underlie human dis-

eases, including fragile X syndrome, frontotemporal dementia, amyotrophic lateral sclerosis, and ataxia with ocular apraxia type 2 (Skourtis-Stathaki and Proudfoot, 2014).

The observation that R-loops form at class switch sequences in activated murine B cells provided the first evidence that R-loops form under physiological conditions (Yu et al., 2003). Our recent work showed that sites of stable DNA:RNA hybrid formation can be profiled genome-wide in human cells using DNA:RNA immunoprecipitation coupled to high-throughput sequencing (DRIP-seq), a method that relies on the unique specificity of the S9.6 monoclonal antibody (Boguslawski et al., 1986) for RNA:DNA hybrids and R-loops independent of the DNA sequence. This revealed that R-loops form normally over unmethylated CpG island (CGI) promoters (Ginno et al., 2012). The observation that R-loops also form at the 3' end of a number of human genes (Ginno et al., 2013) together with evidence that terminal R-loop formation is a key step in transcription termination (Skourtis-Stathaki et al., 2011) suggests that R-loops may possess more general physiological roles (Skourtis-Stathaki and Proudfoot, 2014).

Despite the rising importance of R-loops in pathological and physiological processes, little remains known about their prevalence and distribution in the human genome, their conservation across cell types and species, their mechanisms of formation and turnover, and their possible functions. Here we used a novel iteration of our original method, called DNA:RNA immunoprecipitation followed by cDNA conversion coupled to high-throughput sequencing (DRIPc-seq), to profile R-loops genome-wide at near base-pair resolution and in a strand-specific manner. Our data support the notion that the hybridization of nascent RNA transcripts to template DNA during transcription is a programmed event that associates with a variety of functional biological outputs underlying transcription regulation and chromatin patterning.

RESULTS

High-Resolution, Strand-Specific Profiling of RNA:DNA Hybrids

We developed a high-resolution, strand-specific R-loop profiling method termed DRIPc-seq. This method builds on DRIP, except

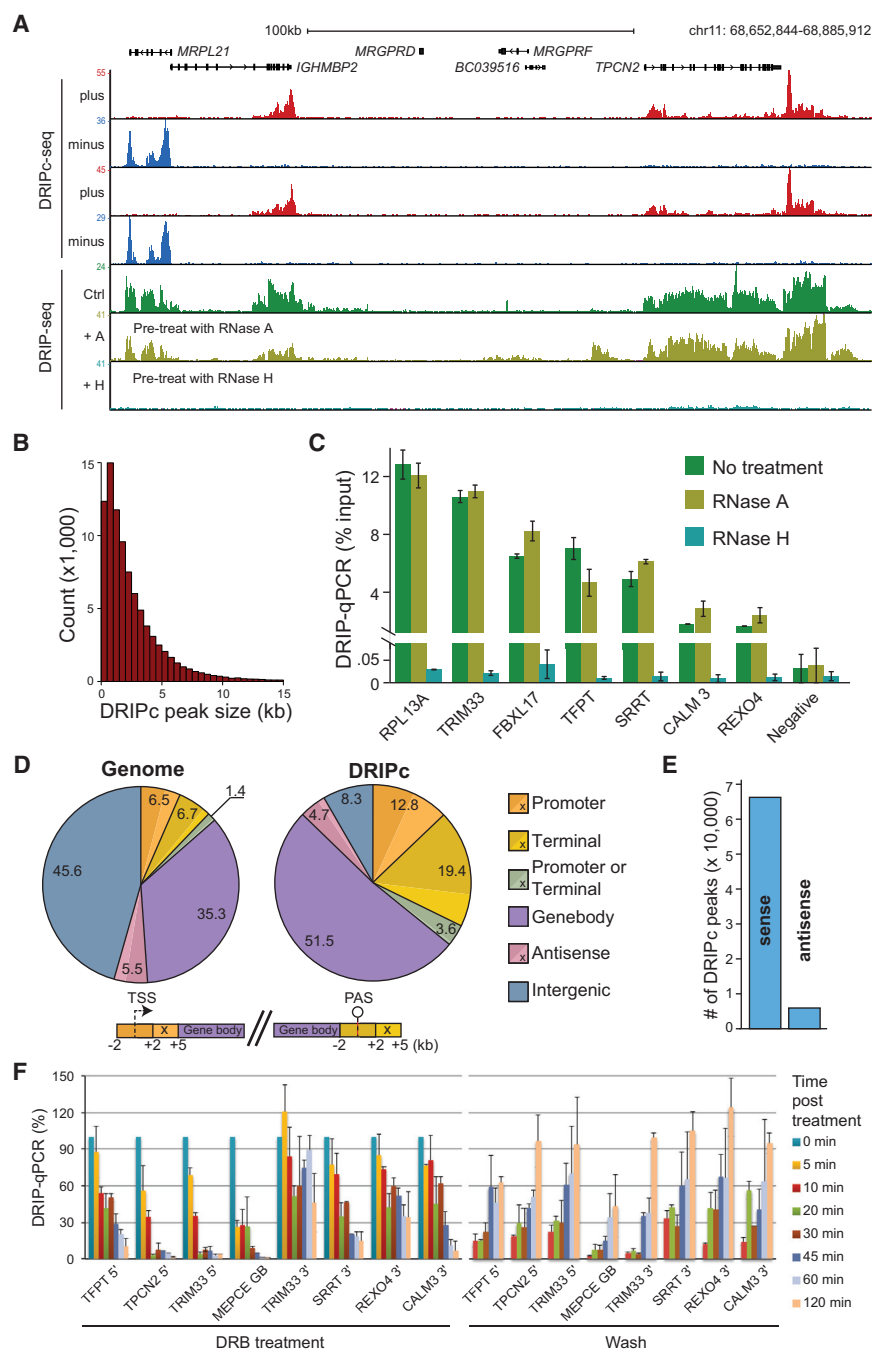


Figure 1. Distribution, Frequency, and Dynamics of RNA:DNA Hybrids

(A) Screenshot of a representative genomic region. DRIPc data are shown in red (+ strand) and blue (– strand) (two independent replicates). DRIP-seq data are shown in green. DRIP-seq data after RNase A and RNase H pre-treatment are shown below (in khaki and teal, respectively).

(B) Distribution of DRIPc peak numbers as a function of peak size.

(C) Bar chart of DRIPc-qPCR (as percent input) for various loci. Each bar is the average of two independent experiments (shown with SE). The effect of RNase A and RNase H pre-treatment on DRIPc-qPCR is shown.

(D) Location analysis of DRIPc peaks (right) compared with expected genomic distribution (left). The various regions are color-coded according to the schematic below. X refers to an extended 3-kb region as shown.

(E) Number of DRIPc peaks in the sense and antisense orientations relative to gene transcription.

(F) R-loops were measured by DRIPc-qPCR at a number of loci through a time course after DRB treatment and wash (5' indicates promoters, 3' indicates terminators). The initial DRIPc-qPCR value for each locus prior to DRB treatment (time zero) was normalized to 100%. Error bars represent the average of two independent experiments with SE.

but suffered from limited (approximately kilobase) resolution, a higher background, and a lack of strand specificity. Importantly, treatment of genomic DNA with Ribonuclease A (RNase A) prior to DRIP resulted in no significant difference in signal (Figure 1A; Figure S1B). By contrast, pre-treatment of genomic DNA with purified human Ribonuclease H2 (RNase H) abolished the R-loop signal. This shows that DRIPc-seq allows the specific recovery of RNA:DNA hybrids. DRIPc-seq was strand-specific and showed a lower background and greater resolution than DRIP-seq (Figure 1A). DRIP-seq and DRIPc-seq (referred to hereafter as DRIP and DRIPc, respectively) were reproducible between biological replicates and in strong agreement with each other (Figures S1B and S1C). The observation that the DRIP signal (which is overwhelmingly RNase H-sensitive and results from direct sequencing of DNA fragments) and the RNA-based DRIPc signal are in strong agreement argues that DRIPc specifically queries RNA strands involved in RNA:DNA interactions and not from residual free RNA. To further establish this, we showed that RNase H treatment after DRIP abrogated our ability to detect a signal after reverse transcription (RT) at all tested loci, including highly transcribed genes (Figure S1D). Likewise, DRIPc sequencing

that S9.6 immunoprecipitated fragments were further digested by DNase I to remove any DNA, and the RNA strands were recovered, reverse-transcribed to cDNA, and subjected to a strand-specific RNA sequencing (RNA-seq) protocol (see [Experimental Procedures](#) for details; Figure S1A). This technique allows the capture of a highly enriched population of RNA molecules involved in RNA:DNA hybrid interactions.

DRIPc-seq was performed on the human embryonic carcinoma Ntera2 cell line. DRIP-seq was performed in parallel for comparison. As expected, DRIP-seq resulted in a robust signal

that S9.6 immunoprecipitated fragments were further digested by DNase I to remove any DNA, and the RNA strands were recovered, reverse-transcribed to cDNA, and subjected to a strand-specific RNA sequencing (RNA-seq) protocol (see [Experimental Procedures](#) for details; Figure S1A). This technique allows the capture of a highly enriched population of RNA molecules involved in RNA:DNA hybrid interactions.

libraries could not be built from RNase H-treated material because not enough material could be recovered (data not shown). By contrast, pre-treatment with RNase A did not affect DRIPc (see below). This, together with the fact that the DRIPc signal showed no enrichment over exons as would be expected from residual mRNA (Figure 1A), established that the method is specific. Finally, omission of the RT step led to a complete lack of signal over tested loci, showing that DNA strands were fully digested by DNase I after DRIP (Figure S1D). DRIPc-seq enables RNA:DNA hybrid profiling at high resolution and with strand specificity in any genome.

Prevalent R-loop Formation over Genic Regions

Based on DRIPc datasets, up to ~150 Mb of DNA distributed among ~70,000 peaks can exist under an R-loop form, representing 5% of the human genome. This is in close agreement with DRIP-based data generated from primary fibroblasts (Lim et al., 2015). Although individual R-loop footprints cannot be measured, population average DRIPc peaks showed a median size of 1.5 kb (Figure 1B). We used DRIP-qPCR to estimate steady-state R-loop formation frequencies. At positive loci, frequencies ranged from ~2%–12% of input depending on the locus, whereas negative loci (intergenic regions and untranscribed genes) ranged from 0.01%–0.1% (Figure 1C). As expected, the DRIP-qPCR signal was insensitive to RNase A and abrogated by RNase H pre-treatment. Thus, DRIP allows the assessment of RNA:DNA hybrid formation with a 100-fold dynamic range.

The vast majority of DRIPc signal mapped onto RNA polymerase II-dependent genes, with 2- and 3-fold over-representation at promoter and terminal regions, respectively (Figure 1D). The largest fraction of R-loops was observed over gene bodies. The intergenic signal was depleted more than 5-fold, and a significant fraction of this signal corresponded to unannotated transcripts and extensions of putative long non-coding RNA genes (Figure S1E). In contrast to yeast (El Hage et al., 2014), little signal could be detected over tRNA genes in human (Figure S1F). Genic R-loop formation was highly strand-specific, with over 90% of the R-loop signal resulting from RNAs transcribed co-linearly with transcription (Figure 1E). This, together with the strong agreement of DRIP and DRIPc signals (Figure 1A; Figure S1C), argues for co-transcriptional structure formation in *cis*. In agreement, the R-loop signal (measured by DRIPc-seq) was positively correlated with gene expression (measured by RNA-seq) (Figure S1G).

Co-transcriptional R-loop Formation Is Dynamic

To further test the dependence of R-loop formation on active transcription, we treated cells with 5,6-dichloro-1- β -D-ribofuranosyl-benzimidazole (DRB), a CDK9 inhibitor that rapidly arrests transcription initiation. R-loops were then measured at time points after treatment by DRIP-qPCR for a range of promoter and terminal loci. At promoters, a rapid drop in R-loop levels followed DRB treatment (Figure 1F, left). Most structures disappeared within 30 min after transcription block, with loci showing an average 10-min half-life. At terminal regions, a clear albeit slower kinetic response was also observed. Such a delayed response is expected given that RNA polymerase II complexes released from the promoter prior to DRB addition must travel

the distance of the gene before R-loops can be resolved. For instance, although R-loops dropped rapidly at the *TRIM33* promoter, *TRIM33* terminator R-loops held steady for 60 min and only began to drop at 120 min. This response is consistent with elongation rates of ~2 kb/min (Jonkers and Lis, 2015) and the fact that *TRIM33* is 118 kb long. R-loops at the 3' ends of *SRRT*, *REXO4*, and *CALM3*, three short genes (9.5–13.6 kb), showed a half-life of about 30 min, which can roughly be accounted for by assuming an ~10 min travel time through the gene body and a 10- to 20-min half-life. Washing DRB 2 hr after treatment led to the progressive reappearance of R-loop structures. At promoters, R-loops started appearing within 10 min post-wash, but steady-state pre-DRB levels were only reached after 2 hr (Figure 1F, right). The kinetics of R-loop re-emergence after DRB was similar at terminator regions of short genes but markedly slower for the longer *TRIM33* gene, as expected. Together, our data support a co-transcriptional origin for most instances of R-loop formation and show that R-loops are dynamically created and resolved.

R-loop Formation over Promoter and Poly Adenylation-Dependent Termination Regions Is a Prevalent Feature of Expressed Human Genes

Promoter R-loop formation was observed for 8,112 genes. R-loop (+) promoters were enriched for CpG island promoters (Figures S2A–S2D), as expected (Ginno et al., 2013). The promoter R-loop signal rose downstream of the transcription start site (TSS) and reached a maximum about 1.5 kb downstream (Figures 2A and 2B), suggesting that RNA:DNA hybrid formation often spans the promoter-proximal part of the first intron. Guanine cytosine (GC) skew, which measures strand asymmetry in the distribution of guanines and cytosines, was associated with promoter R-loop formation (Figures 2A and 2B), consistent with the higher thermodynamic stability of RNA:DNA hybrids carrying G-rich RNA strands (Ratmeyer et al., 1994).

Terminal R-loop formation was observed for 9,320 genes. The terminal R-loop signal was broad and peaked just before the polyadenylation site (PAS) (Figure 2C). Colliding genes showed a strong terminal signal emanating from each gene (Figure 2D). Terminal R-loops were accompanied by only modest GC skew transitions around the PAS. Most genes with 3' R-loops did not associate with significant GC skew, as measured by the SkewR algorithm (Ginno et al., 2013). Interestingly, terminal R-loop formation was specifically observed for polyadenylation [poly(A)]-dependent genes. Highly expressed genes, such as histone genes, certain long non-coding RNAs, and many RNA polymerase III transcripts that do not undergo poly(A)-dependent processing, did not show evidence of terminal R-loops.

Gene Body R-loop Formation Is Prevalent

RNA-DNA interactions in gene bodies were frequent (Figure 2E). In general, the cumulative length of R-loop peaks in a gene was proportional to that gene's length (Figure S2E), indicating that transcription elongation through longer genes may be more R-loop-prone. For nearly two-thirds of genes, the R-loop signal occupied less than 30% of the gene body sequence space (Figure 2E). However, the R-loop signal extended over 50%–100% of the gene body for a quarter of genes. These “sticky” genes

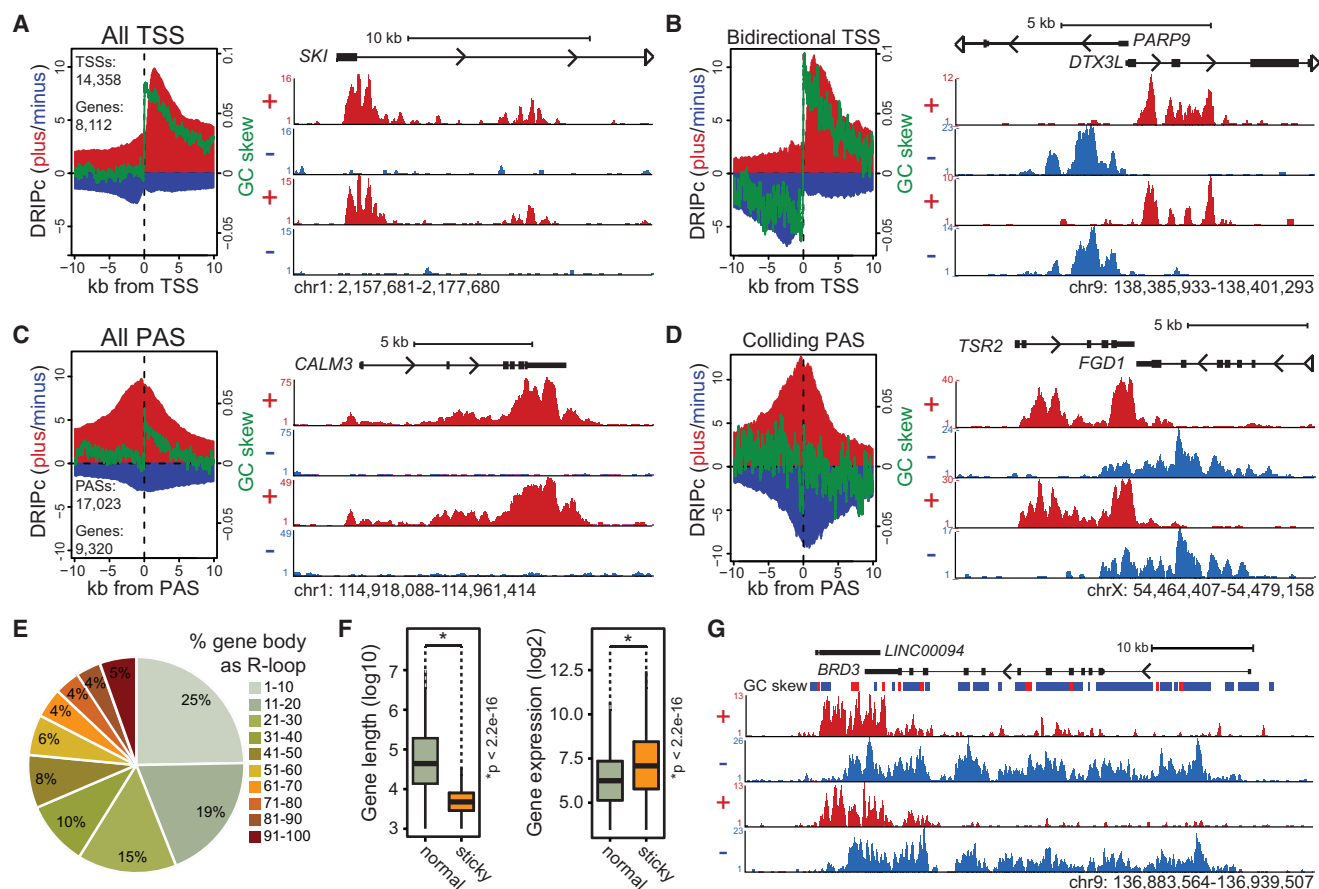


Figure 2. Prevalent R-loop Formation

(A–D) Left: metaplot of DRIPc signal centered on the TSS (A and B) or PAS (C and D). Red, + strand signal; blue, – strand signal. All genes are oriented to the right. Green, average GC skew for corresponding genes. Right: representative screen shots for each gene class.

(A) All R-loop-forming TSSs.

(B) Bidirectional R-loop-forming TSSs.

(C) All R-loop-forming terminators.

(D) R-loop-forming colliding terminators.

(E) Pie chart indicating the fraction of genes (in percent) carrying a gene body R-loop signal (in percent gene body covered by signal grouped by decile).

(F) Length and expression characteristics of sticky genes versus normal R-loop-forming genes.

(G) Representative screenshot of a sticky gene, *BRD3*. GC skew is indicated as red (positive skew) and blue (negative skew) blocks.

tended to be shorter and/or highly transcribed (Figure 2F; Figure S2F), and some were accompanied by extensive gene body GC skew (Figure 2G). The gene body signal was highly reproducible (Figure 2G and below), arguing that RNA–DNA interactions during elongation occur at specific positions along genes.

R-loop Formation Is Conserved within Human Cell Types and across Species

To address R-loop conservation, we performed DRIP-seq in human K562 (erythroleukemia) cells and in mouse E14 embryonic stem cells (ESCs) and NIH 3T3 cells (embryonic fibroblasts). Our recent R-loop map from human primary fibroblasts (Lim et al., 2015) was also analyzed, and DRIPc-seq was performed on NIH 3T3 after RNase A pre-treatment. R-loop formation was prevalent, biased over genes, and enriched over promoters

and terminators in both human and mouse cells (Figure S3A). Importantly, R-loop formation was conserved across cell types and species. For instance, regions in human chromosome 14 that are syntenic to mouse chromosome 12 and 14 displayed a high conservation of the DRIP signal (Figure 3A). This conservation was reflected by high Pearson correlation when comparing DRIP signals over orthologous genes (Figure 3B, left) or over 1-kb syntenic genomic tiles (Figure 3B, right). In most cases, differences in R-loop formation over orthologous genes could be explained by differences in gene expression (Figures S3B and S3C). We also analyzed how conservation of the signal varied over genic sub-regions and determined that promoter and terminal regions showed the highest agreement, ranging from 40%–60% when comparing within species and to 25%–40% when comparing between species (Figure 3C). A slightly lower but significant conservation was observed throughout gene body

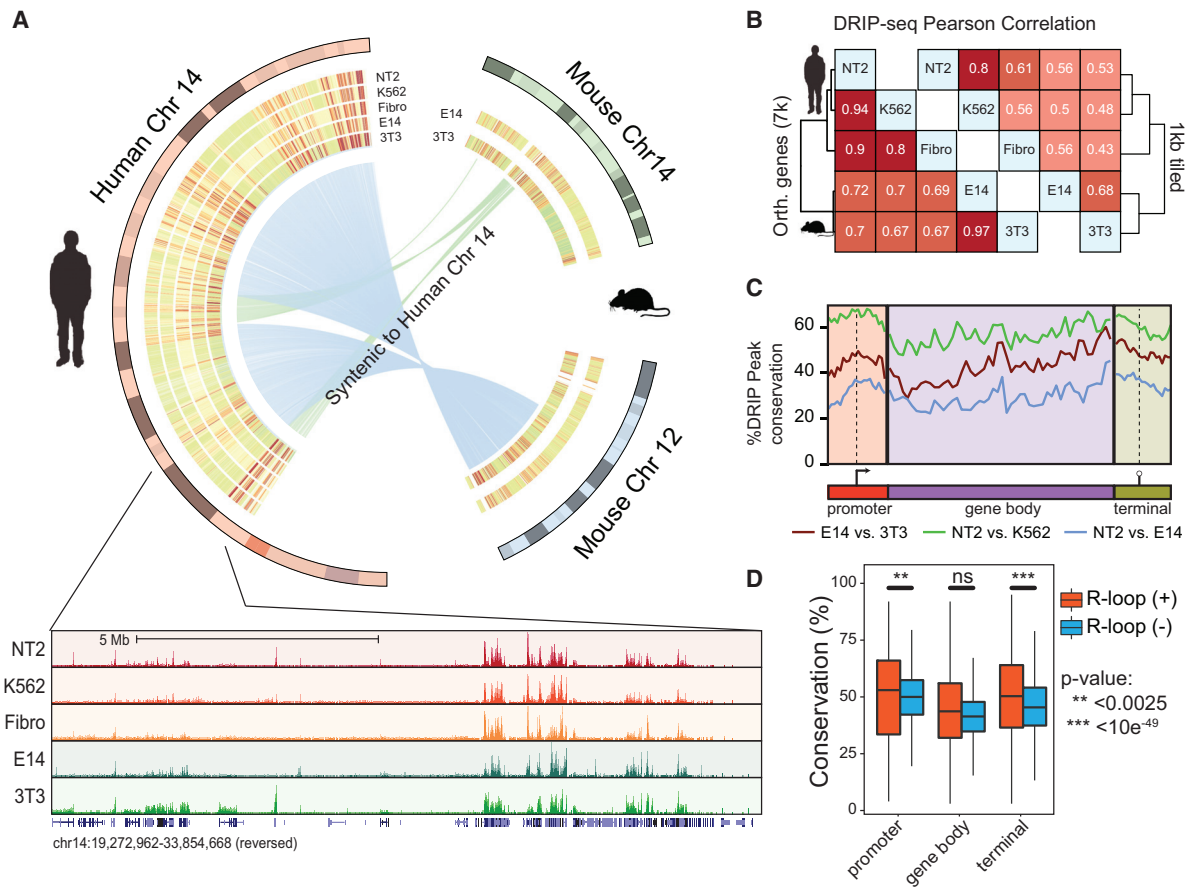


Figure 3. R-loop Conservation

(A) Top: circos plot depicting the DRIP signal (red, high; green, low) along human chromosome 14 in 1-Mb bins for human Ntera2 (NT2), K562, and fibroblasts (Fibro) as well as murine E14 and NIH 3T3 (3T3) cells. Regions from mouse chromosomes 14 and 12 syntenic to human chromosome 14 are connected by ribbons. Bottom: enlarged region; the genes are indicated below.

(B) Human versus mouse Pearson correlation of the DRIP signal over orthologous genes (left) and 1-kb genomic tiles (right).

(C) Conservation of the R-loop signal between mouse and human cell lines broken down by gene parts.

(D) Sequence conservation over genic regions for R-loop (+) (orange) and (–) (blue) loci matched for expression. p Values (Wilcoxon Mann-Whitney one-tailed test) indicate significance.

regions. At higher resolution using DRIPc datasets in human and mouse, conservation was also strong, in particular over promoters and terminators (Figures S3D and S3E). To test whether DNA sequence conservation underlies R-loop conservation, we compared DNA sequences between orthologous R-loop (+) and (–) regions. Conservation was significantly higher for terminal and promoter regions but not for gene bodies (Figure 3D). Together, this establishes that R-loops often form at specific conserved loci in the human and mouse genomes.

R-loop Regions Exhibit Increased Overall Chromatin Accessibility

Our DRIPc maps enabled us to determine with unprecedented resolution whether R-loop formation associates with particular chromatin states. For this, we obtained the Encyclopedia of DNA Elements (ENCODE) and other datasets representing a range of chromatin marks, including DNase hypersensitivity, micrococcal nuclease cleavage, a number of histone modifica-

tions, and various chromatin binding factors (Table S1). To determine whether a chromatin trait was enriched or depleted over R-loop loci, we measured the intensity of that trait over DRIPc peaks and compared it with the intensity for a shuffled set of non-R-loop-forming peaks. To ensure adequate comparisons, we treated promoter, gene body, and terminal loci separately. Shuffling was conducted within each genic sub-region while strictly matching peak size and location relative to TSS and PAS and matching gene expression levels and gene lengths (Supplemental Experimental Procedures). This analysis allowed us to identify chromatin signals that distinguish R-loop-positive (+) and negative (–) loci without interference from location or expression.

All R-loop (+) regions showed significantly higher DNase accessibility compared with R-loop (–) loci (Figure 4). For promoters, increased accessibility was most marked 1–2 kb upstream of the R-loop peak, corresponding to the TSS. DNase accessibility was maximal at the center of the R-loop peak for

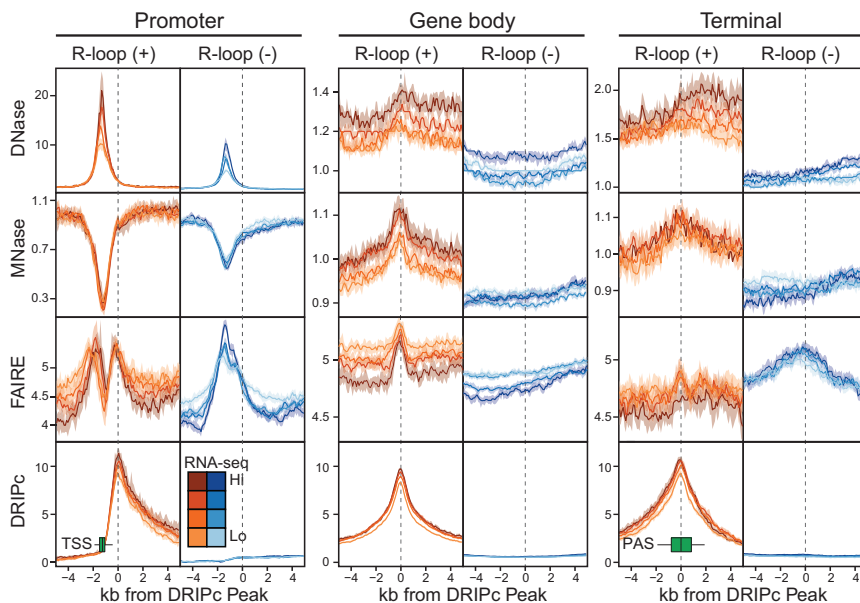


Figure 4. R-loop-Forming Regions Are Associated with Open Chromatin and High RNA Polymerase Occupancy

All metaplots are centered over R-loop peaks. R-loop (+) and R-loop (–) represent loci with and without R-loops, respectively. R-loop (+) and (–) genes were matched for expression and grouped into four expression quartiles (see inset in the DRIPc metaplot). Metaplots of the DNase, MNase, FAIRE, and DRIPc signals are arranged from top to bottom and further broken down by promoter, gene body, and terminal regions from left to right. The horizontal boxplots at the bottom of the promoter and terminal columns indicate the positions of the TSS and transcription termination site (TTS), respectively. Lines indicate the median value of the signal at each position; the 95% confidence interval is shaded.

gene bodies and was broadly distributed around R-loop (+) terminators. Patterns of MNase cleavage confirmed the presence of a hyper-accessible region over the TSS of R-loop (+) promoters (Figure 4). However, R-loop (+) regions in gene bodies and terminators were less prone to MNase cleavage despite higher DNase accessibility. This pattern could reflect a reduced MNase cleavage efficiency at R-loop structures or the possibility that R-loops favor the deposition of labile nucleosomes and/or are bound by non-histone proteins. Formaldehyde-assisted isolation of regulatory elements sequencing (FAIRE-seq) is a third method by which the accessibility of chromatin can be measured (Giresi et al., 2007). R-loop (+) promoters were characterized by a clear second peak of FAIRE-seq signal overlapping with the R-loop peak. This second TSS-distal peak was conspicuously absent in R-loop (–) promoters (Figure 4), suggesting that R-loop formation associates with increased chromatin accessibility. Gene body R-loop peaks also coincided with a clear FAIRE-seq peak compared with matched R-loop (–) controls. At terminal regions, however, R-loop (+) regions showed lower FAIRE accessibility than matched R-loop (–) terminators, suggesting a mixed chromatin architecture characterized by DNase hyper-accessibility and lower MNase and FAIRE sensitivity.

R-loop (+) Promoters Show Enhanced Epigenomic Signatures Characteristic of Open, Active Promoters

Histone 3 lysine 4 (H3K4) di- and tri-methylation, along with H3K9 and H3K27 acetylation, were significantly enriched over promoter R-loops (Figures 5A and 5B). The levels of these four histone marks were ~2-fold higher for R-loop (+) promoters compared with matched R-loop (–) controls. As noted for DNase hypersensitivity, increased deposition was observed over TSSs (Figure 5B). H3K4 mono-methylation was also significantly enriched, but the enrichment was directly over the R-loop peak as well as upstream of the TSS (Figures 5B and 5C). A modest but significant enrichment of the histone variant H3.3 was

observed around TSSs of R-loop (+) promoters, along with a significant enhanced recruitment of the transcription elongation mark H3K36me3 into gene bodies (Figures 5B and 5C). Increased levels of H3K36 tri-methylation (H3K36me3) coincided with an elevated density of active RNA polymerase (RNAP) complexes as measured by precise maps of polymerase (PRO)-seq (Kwak et al., 2013; Figure 5B; Figure S4A). Overall, R-loop (+) promoters displayed chromatin hyper-accessibility, hyper-acetylation, hyper-methylation of H3K4, and enhanced H3K36me3 deposition compared with expression-matched R-loop (–) promoters. In addition, we observed depletion of the heterochromatic mark H3K9me3 (Figure 5B) and enhanced protection against DNA methylation at R-loop (+) promoters as well as promoter-proximal R-loop peaks (Figure 5D; Figures S4B and S4C), indicative of a hyper-protected state against epigenetic silencing. Analysis of chromatin hidden Markov model (chromHMM) states (Ernst and Kellis, 2012) confirmed significant associations with strong promoters, enhancers, and transcriptional transition states accompanied by a depletion of the heterochromatic state (Figure S4D).

The levels of multiple chromatin binding factors also responded to R-loop formation. RBBP5, a core component of the complex proteins associated with Set1 (COMPASS) H3K4 methyltransferase complexes (Shilatifard, 2012) was significantly enriched at R-loop (+) promoters (Figures 5A and 5E). PAF1, a subunit of the RNAP-interacting PAF1 complex (PAF1C), was also enriched (Figures 5A and 5E) together with the p300 acetyltransferase (Figure 5A; Figure S4E). SIN3A, SAP30, and HDAC2, components of the SIN3 complex, were similarly enriched around the TSS of R-loop (+) genes (Figures 5A and 5E). KDM4A and PHF8, two histone demethylases targeting H3K9/H3K36 and H3K9/H4K20, respectively (Klose et al., 2006; Liu et al., 2010) were also highly enriched, consistent with the observed reduction of H3K9me3 levels. ZNF274, a Krüppel associated box (KRAB) zinc finger protein recruited to H3K9me3-marked regions (Fietze et al., 2010), was conversely depleted (Figures 5A and 5E). Finally, we observed increased recruitment of the EZH2 methyltransferase and increased

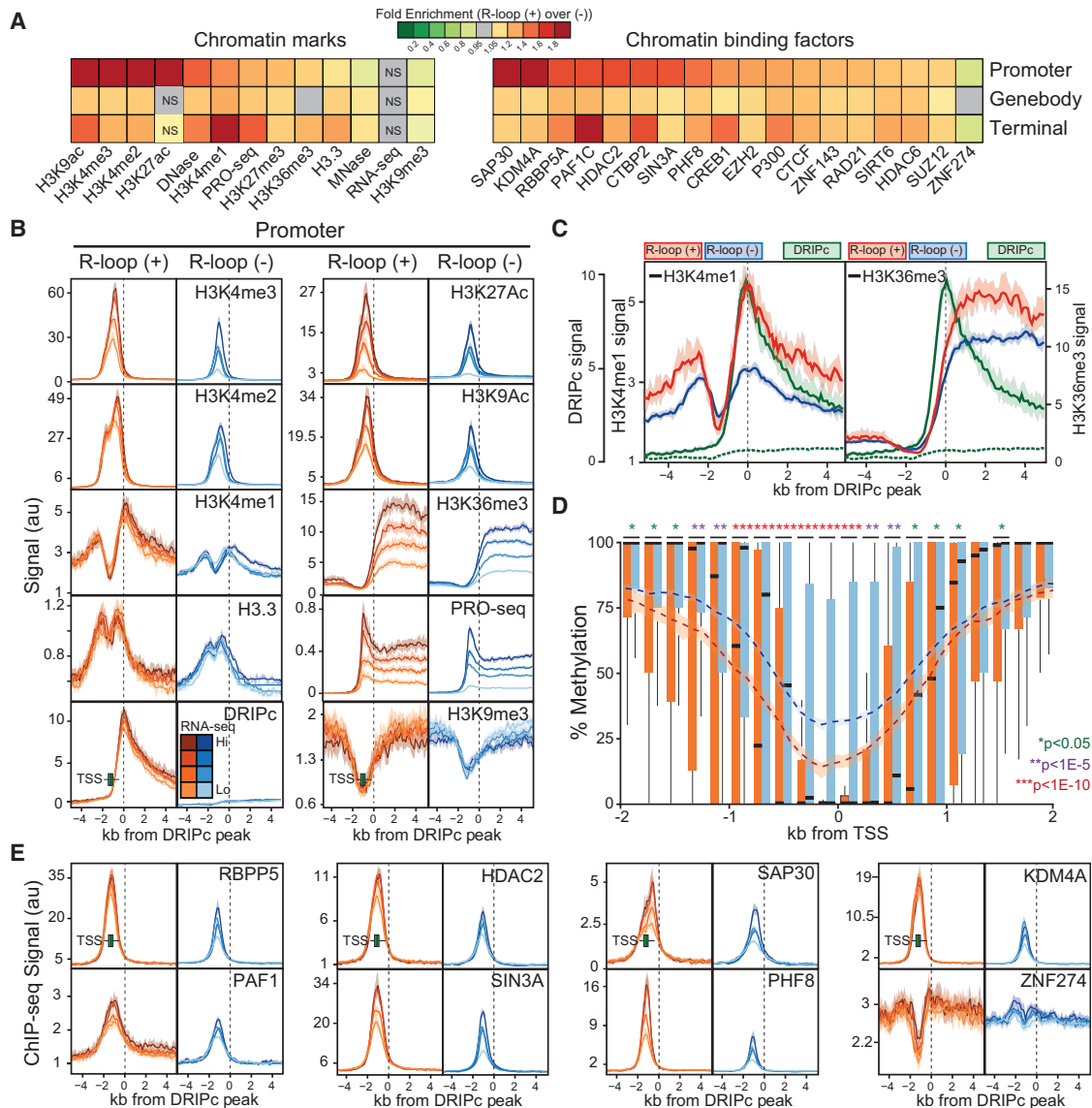


Figure 5. R-loop-Forming Promoters Associate with Specific Histone Marks, Chromatin Binding Factors, and DNA Hypomethylation

(A) Heatmap showing chromatin marks and chromatin binding factor enrichment or depletion over R-loop regions (promoter, gene body, and terminal) relative to shuffled regions. Grey indicates an effect size of less than 5%. NS, not significant. Other differences show $p < 0.05$ (Monte Carlo).

(B) Metaplots of the ChIP-seq signals for H3K4me1/2/3, H3K9ac, H3K27ac, H3K36me3, H3K9me3, H3.3, and PRO-seq with the DRIPc signal at promoters. The colors and organization are as described for Figure 4.

(C) Overlay of H3K4me1 (left) and H3K36me3 (right) ChIP-seq over the DRIPc signal. The ChIP-seq signal is broken down between R-loop (+) genes (red) and expression-matched R-loop (-) genes (blue). Highly expressed genes (Q4) are shown. Green lines represent the DRIPc signal for R-loop (+) (solid line) and R-loop (-) (dotted line) Q4 genes. 95% confidence intervals are shaded.

(D) Boxplot of DNA methylation over a TSS-centered, 4-kb region (low expression quartile, Q1, is shown). Each boxplot pair corresponds to a 200-bp window over R-loop (+) (orange) and expression-matched R-loop (-) (blue) promoters. The median is shown by a black line. For each window, p values indicate the significance of any difference between R-loop (+) and (-). Dotted lines indicate average DNA methylation values over the entire region.

(E) Metaplots of the ChIP-seq signals for RBPP5, PAF1, HDAC2, SIN3A, SAP30, PHF8, KDM4A, and ZNF274 with the DRIPc signal.

deposition of the polycomb mark H3K27me3 over R-loop (+) promoters in a manner inversely proportional to gene expression (Figure S4E). Overall, our data suggest that promoter R-loop formation is accompanied by the enhanced recruitment or depletion of specific chromatin binding factors, resulting in unique epigenomic signatures.

Terminal R-loops Show an Enhancer and Insulator Chromatin State

R-loop (+) terminator regions showed the strongest enrichment for the enhancer marks H3K4 mono-methylation (H3K4me1) and p300 (Heintzman et al., 2007; Figures 5A and 6A). In agreement, R-loop (+) terminators were two to three times more likely

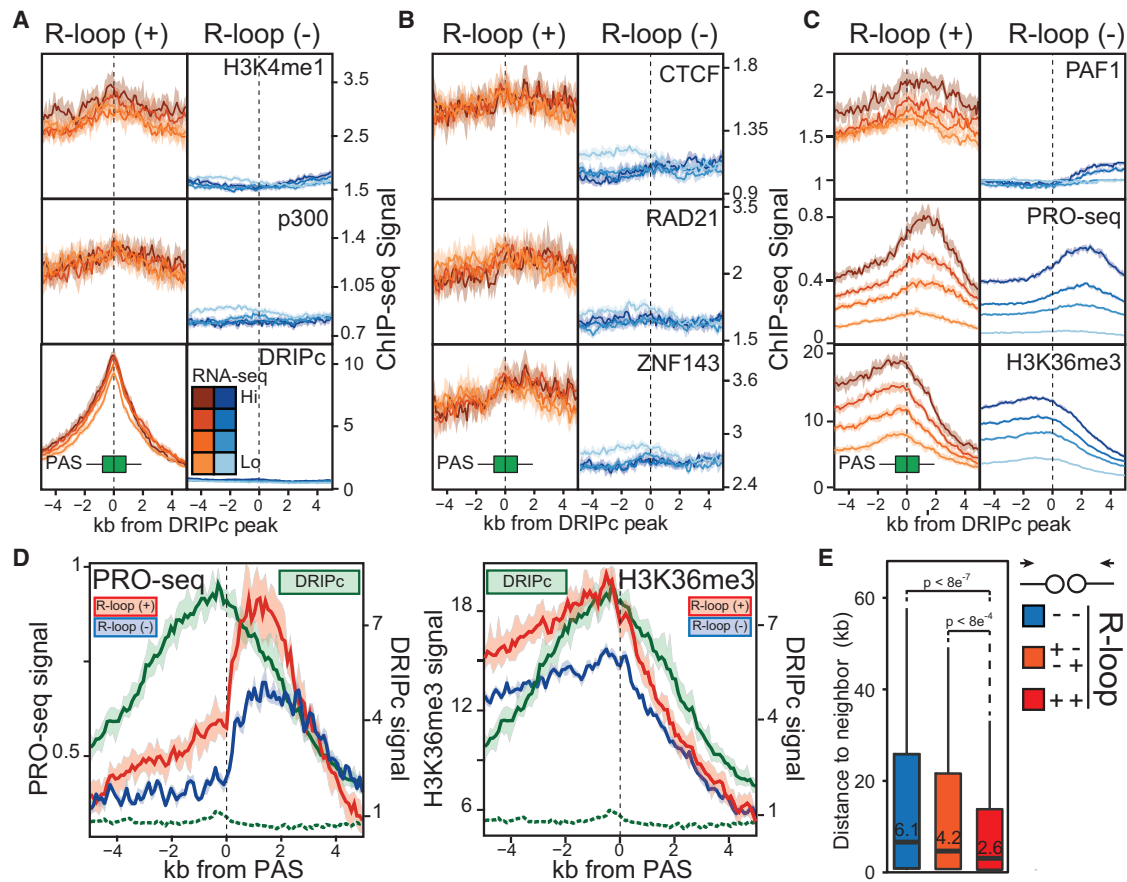


Figure 6. R-loop-Forming Terminators Are Associated with Enhancer and Insulator-like State and Show Characteristics of Transcription Terminators

(A) Metaplots of the ChIP-seq signal for H3K4me1 and p300 with the DRIPc signal over terminal R-loop regions. The plots are centered on terminal R-loop peaks. The color code is as described for Figure 4.

(B) Same as (A), except CTCF, RAD21, and ZNF143 were analyzed.

(C) Same as (A), except PAF1, PRO-seq, and H3K36me3 were analyzed.

(D) Overlay of PRO-seq (left) and H3K36me3 ChIP-seq (right) over the highly expressed (Q4) terminal DRIPc signal for R-loop (+) (red) and R-loop (-) genes. See Figure 5C for color codes.

(E) Nearest neighbor gene distances for expressed (>Q1) colliding gene neighbors where none, one, or two of the neighbors form terminal R-loops. p Values were determined by a Wilcoxon Mann-Whitney test.

to overlap with annotated enhancers (Figure S5A), and the “strong enhancer” state was the most significantly enriched chromHMM state over terminal R-loops (Figure S4D). Whether such enhancers are active is unclear given the low H3K27 acetylation (H3K27Ac) levels and the slight enrichment of H3K27me3 (Figure S5B). Overall, as judged by increased DNase accessibility (Figure 4), H3K4me1 deposition, p300 recruitment, and enhancer annotations, terminal R-loops associate with an enhancer state. Terminal R-loop formation was also associated with increased levels of CTCF, a zinc finger DNA binding protein with insulator function that serves as a major organizer of chromatin loops (Bell et al., 1999; Figure 6B). A significant enrichment for ZNF143 and RAD21, components of the cohesin complex that often co-localize with CTCF (Parelho et al., 2008), was also observed. ChromHMM confirmed that the “insulator” state was specifically enriched at terminal R-loop regions, not promoters (Figure S4D). Recent evidence suggests that terminal

R-loops associate with a condensed, H3K9 di-methylation (H3K9me2)/3-marked state (Skourti-Stathaki et al., 2014). Epigenomic profiling of human R-loop (+) termini, however, indicated that these regions are DNase hyper-accessible (Figure 4) and depleted for H3K9me3 (Figure 5A; Figure S5B) and heterochromatic traits (Figure S4D). Analysis of H3K9me2 and other datasets available for the mouse broadly confirmed that terminal R-loops are open and enriched for H3K4me1 and H3K36me3 (Figure S5C) compared with matched R-loop (-) termini. However, no global enrichment for H3K9me2/3 could be observed, although this association may hold at specific genes.

Terminal R-loop Regions Show Characteristics of Transcription Terminators

PAF1 was also highly enriched at the 3' ends of R-loop-forming genes (Figure 6C). This observation is significant given that PAF1C interacts with 3' end processing factors such as the

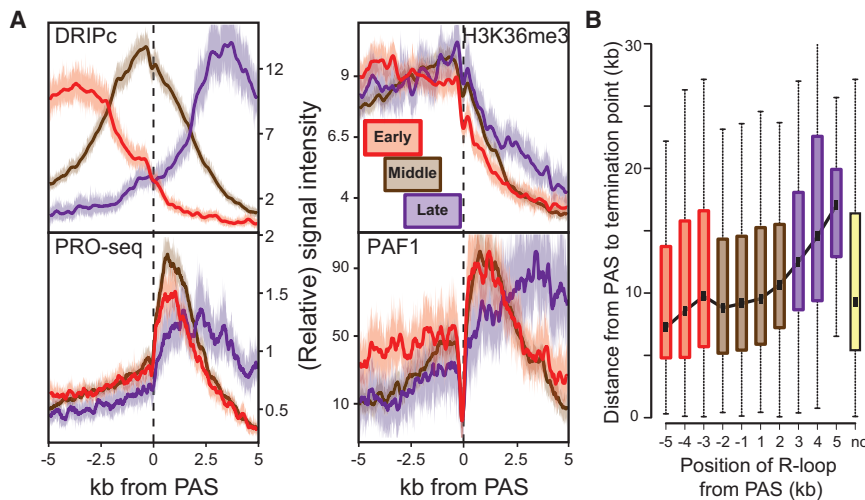


Figure 7. The Position of the R-loop Signal Correlates with the Position of Transcription Termination

(A) Metaplots of DRIPc, PRO-seq, H3K36me3, and PAF1 signals over terminal R-loop (+) genes (Q4). In each case, the signal is centered on the PAS and broken down between early ($n = 2,095$), middle ($n = 4,280$), and late-forming ($n = 1,126$) R-loop (+) genes. Lines represent median values. 95% confidence intervals are shown. For PAF1, signal intensities were normalized and reported as percent maximal signal to better define the positional features.

(B) Distribution of distances between the PAS site and the transcription termination points of individual genes as a function of R-loop signal position arranged in 1-kb bins around the PAS. The boxplots include all genes, and a median trend line is shown. Data were calculated from PRO-seq data (K562). Genes with no terminal R-loops are shown at the right.

cleavage and polyadenylation specificity factors (CPSFs) (Tomson and Arndt, 2013) and that R-loops have been linked to transcription termination (Skourti-Stathaki et al., 2011). To further analyze the association between terminal R-loops and termination pathways, we measured the density of active RNAP taking advantage of PRO-seq datasets. R-loop (+) termini showed a significantly greater RNAP density compared with expression-matched R-loop (−) termini (Figure 6C). RNAP density for R-loop (+) loci started rising as early as 4 kb before the PAS, coinciding with an increased R-loop signal (Figure 6D). In contrast, RNAP density for R-loop (−) loci remained flat before the PAS. At the PAS, a sharp rise of RNAP density was observed, most likely reflecting termination-associated stalling of the RNAP complex. RNAP density post-PAS peaked higher and earlier and started to return back down sooner and at a faster rate for R-loop (+) than for R-loop (−) termini. As a result, RNAP density returned toward background levels earlier for R-loop-forming termini, consistent with the notion that terminal R-loops favor transcription termination. In agreement, H3K36me3 levels rose up to the R-loop peak and rapidly decreased past that point for R-loop (+) genes (Figure 6C). By contrast, R-loop (−) genes showed lower H3K36me3 levels and a decreased rate of signal decay (Figures 6C and 6D). Thus, the transcription complex rapidly dissociates over terminal R-loops, a characteristic of transcription terminators. In general, genes with R-loop (+) terminators tended to be closer to their nearest 3′ neighbor compared with expression-matched terminal R-loop (−) genes (median distances 12 kb versus 17 kb, respectively; p value $< 2e^{-5}$). This trend was true for co-directional gene neighbors and particularly noticeable for colliding genes where the presence of one or two R-loop (+) terminator(s) was correlated with shorter and shorter intergene distances (Figure 6E). Vice versa, colliding genes with short inter-gene distances were enriched for R-loop (+) terminators compared with colliding genes with long intergene distances for which they were depleted (Figure S6A). This suggests that terminal R-loop formation defines a class of transcription terminators that are enriched for genes separated by shorter intergenic distances.

The Position of Terminal R-loops Influences the Position of Transcription Termination

The broadness of the R-loop signal observed around the PAS in metaplots suggests that early, middle, and late R-loop-forming loci may exist. To test this, we systematically determined the center of gravity of the R-loop signal for each annotated terminal peak and assigned the position of that point to 1-kb bins distributed $\geq \pm 5$ kb around the PAS. Bins ranging from -5 to -2 kb were called “early,” bins from -2 to $+2$ kb were called “middle,” and bins from $+2$ to $+5$ kb were deemed “late” (Figure 7A). This allowed us to determine whether the position of R-loops relative to the PAS influences termination, as measured through the patterns of accumulation and clearance of RNAP, H3K36me3, and PAF1. RNAP density clearly responded to R-loop position, as measured by PRO-seq, native elongating transcript (NET)-seq, and RNAP chromatin immunoprecipitation sequencing (ChIP-seq) data (Figure 7A; Figure S6A). Genes with early R-loops showed the earliest accumulation of RNAP after the PAS, followed by the earliest return of RNAP density back to background levels. By contrast, genes with late R-loops showed a profound delay in RNAP accumulation and a much reduced rate of RNAP clearance from the chromatin template (Figure 7A; Figure S6A). Similar patterns were observed for H3K36me3. Finally, early and late R-loop-forming terminators showed predominant PAF1 levels over early and late regions, respectively, with middle R-loop-forming genes showing the highest PAF1 recruitment around the PAS. This suggests that the position of the R-loop signal relative to the PAS influences the transcription termination process.

To further investigate the relationship between terminal R-loop position and transcription termination, we annotated the distal positions at which transcription terminates for individual genes using available PRO-seq (Kwak et al., 2013) and NET-seq (Mayer et al., 2015) datasets. Annotation was performed using custom hidden Markov models (see Experimental Procedures for details; Figure S6B). The resulting data revealed a strong positive relationship between the position of the R-loop signal and distal termination points (Figure 7B). Genes with early

R-loops terminated earliest (~ 7 kb downstream of the PAS according to PRO-seq data), whereas genes with late R-loops terminated latest (~ 17 kb downstream). By contrast, genes without terminal R-loops terminated at a median value of 9 kb downstream of the PAS. Slightly shorter distances were obtained from NET-seq data, but identical trends prevailed (Figure S6C). The effect of R-loop position on termination was observed regardless of gene expression, with highly expressed genes terminating further away from the PAS than less expressed genes (Figure S6D). Together, this suggests that the position at which the RNAP is released from chromatin is influenced by the presence and the position of terminal R-loops.

DISCUSSION

Prevalent and Dynamic R-loop Formation in Mammalian Genomes

Our results establish co-transcriptional R-loop formation as an important feature of RNA polymerase II-dependent transcription in mammalian genomes. Unlike previously thought, RNA-DNA interactions are not rare under physiological conditions (Figures 1 and 3). Furthermore, R-loop formation does not result from accidental entanglements of the nascent RNA transcript (Figure 3). CGI promoters and terminator regions of poly(A)-dependent genes represent universal hotspots of R-loop formation. At CGI regions, positive GC skew supports structure formation by ensuring the synthesis of G-rich nascent RNAs that can stably associate with the template DNA strand (Figure 2). Terminal R-loop formation, by contrast, was not as highly associated with GC skew, suggesting that R-loop formation mechanisms may differ between gene ends. Importantly, R-loops formed during transcription are dynamically resolved with a half-life of ~ 10 – 20 min (Figure 1). Such turnover may enable cells to mitigate the negative effects of R-loop formation on genomic stability (Aguilera and García-Muse 2012; Hamperl and Cimprich 2014). The ~ 10 -min half-life measured at promoters is also in close agreement with estimates of the paused promoter-proximal RNAP half-life (~ 7 min) (Jonkers et al., 2014). This suggests that R-loop formation may be compatible with the normal dynamics of the transcription cycle. Efficient R-loop resolution may also allow cells to limit the steady-state frequency of R-loop formation. In agreement, R-loop frequency estimates range from $\sim 2\%$ – 15% of input (Figure 1B). Thus, although a surprisingly large portion of mammalian genomes is R-loop-prone, R-loop formation is regulated in terms of frequency and residence time. Overall, our data suggest that the retention of nascent RNAs at the transcription site is a conserved, prevalent, and dynamic feature of the mammalian chromatin.

R-loops Favor an Open Chromatin State under Physiological Conditions

Regardless of location, R-loops associate with DNase I hyper-accessibility (Figure 4). This is consistent with reports that RNA:DNA hybrids prevent nucleosome deposition in vitro (Dunn and Griffith, 1980) and adopt a more rigid A form-like conformation (Noy et al., 2005). Thus, R-loops may interfere with nucleosome re-deposition behind the advancing RNAP, thereby favoring an open state. Evidence that R-loops promote

chromatin decondensation and lower nucleosome occupancy (Powell et al., 2013) and that, conversely, R-loop destabilization causes chromatin compaction (Boque-Sastre et al., 2015) further supports this view. The association of R-loops with H3S10P and with chromosome condensation reported in R-loop-accumulating mutants (Castellano-Pozo et al., 2013) may denote possible pathological consequences of R-loop formation. Careful measurements of R-loop distribution, frequencies, and turnover rates under various conditions are now possible and should elucidate how R-loops differentially associate with chromatin states.

R-loop Formation Favors Specific Epigenomic Signatures

At promoters, R-loops associate with an open, H3K4 hyper-methylated, hyper-acetylated, and hyper-protected state characteristic of strong CGI promoters. In agreement, a recent report showed that R-loops favor a hyper-acetylated state (Chen et al., 2015). Whether R-loops play a causal role in setting this active chromatin state or follow its establishment is a key question. Our work shows that the presence or absence of R-loops affects the levels of multiple chromatin marks and factors anchored around the TSS region of R-loop (+) promoters located 1–2 kb upstream of the R-loop peak itself (Figures 4 and 5). This is most easily interpreted to suggest that the architecture of these promoters is favorable for R-loop formation. However, some R-loop-responsive chromatin features, including H3K4me1, H3K36me3, RNAP density, and promoter FAIRE-seq, coincide with the location of R-loop peaks. This raises the possibility that R-loops may affect the chromatin state in specific instances. Recent studies support the possibility that R-loops may play a causal role in chromatin patterning. R-loop destabilization at the human *VIM* promoter causes a shift from an active, open, DNA-hypomethylated state to a silent, closed, and methylated state (Boque-Sastre et al., 2015). In mouse ESCs, R-loops were suggested to recruit the Tip60-p400 acetylase complex to promoters (Chen et al., 2015). It is possible that R-loops are directly recognized by proteins with a role in epigenetic patterning, thereby connecting the formation of nucleic acid structures to the establishment of specific chromatin states. Proteins with a capacity for ssDNA and/or RNA:DNA hybrid binding are attractive candidates as possible readers and effectors of R-loop formation. H3K4 methyltransferases of the SET1/COMPASS family were shown to bind to ssDNA and negatively supercoiled DNA in vitro (Krajewski et al., 2005). Here we observed that H3K4me1 was enriched over the peak of the R-loop at promoters and terminators (Figures 5 and 6) and that H3K4me2/3 was enriched around the TSSs of R-loop (+) promoters (Figure 5). Processive H3K4 methylation by COMPASS requires PAF1C (Shilatifard, 2012). We show here that PAF1 is enriched at R-loop (+) promoters and 3' ends (Figures 5, 6, and 7). Interestingly, the RTF1 subunit of PAF1C was reported to bind to ssDNA (de Jong et al., 2008). Furthermore, deletion of the *Leo1* and *Cdc73* PAF1C subunits leads to genomic instability in yeast, and this instability can be corrected by RNase H1 overexpression (Wahba et al., 2011). PAF1C is also required for H3K36 trimethylation by SETD2 (Chu et al., 2007), and we show here that R-loop (+) regions show significantly increased

H3K36me3 deposition at both gene ends. Interestingly, although R-loops tend to generally associate with an open, H3K4/H3K36-methylated state, R-loop (+) promoters and terminators also take on distinct states. In particular, the enhancer and insulator states are dominant or specific for terminal regions only (Figures 5 and 6). This suggests that R-loop formation may influence chromatin states in a context-dependent manner.

R-loop Formation Defines a Class of Transcription Terminators

Our data support the view that terminal R-loops are implicated in transcription termination (Figure 6). Furthermore, the position at which transcription terminated was correlated to the position at which R-loops formed (Figure 7). R-loop formation therefore appears to be a conserved hallmark of a broad class of transcription terminators. The mechanistic roles of R-loops in the termination process remain to be fully established. It is possible that the intrinsic ability of R-loops to stall transcription (Belotserkovskii and Hanawalt, 2011) represents an initial pause signal, as suggested previously (Skourti-Stathaki et al., 2011). The formation of an enhancer-like chromatin state (Figure 6) may account for the observation of non-coding RNA transcripts initiating from the distal ends of mammalian genes, particularly those with short intergenic distances (Carninci et al., 2005), through the production of enhancer RNAs. R-loop-induced antisense transcription over the 3' ends of genes has been linked to transcription termination (Skourti-Stathaki et al., 2014). The increased formation of insulator regions and/or gene loops by R-loop (+) terminators via CTCF and cohesin complex recruitment may also contribute to termination (Grzechnik et al., 2014). Finally, it is possible that specific termination proteins are recruited to R-loop terminators. PAF1C is an interesting candidate for this role given its role in termination and 3' end processing and its strong enrichment at R-loop (+) terminators (Figure 6). Consistent reports of R-loop-mediated instability in CPSF and other 3' end-processing mutants (Gómez-González et al., 2011; Stirling et al., 2012; Wahba et al., 2011) further consolidate the links between R-loop formation and 3' end processing.

EXPERIMENTAL PROCEDURES

DRIP-Seq and DRIPc-Seq

DRIP was performed essentially as described previously (Ginno et al., 2012; Supplemental Experimental Procedures). After DRIP, the eluted DNA was treated with 4 U of DNase I (New England Biolabs) for 35 min at 37°C to degrade all DNA. The RNA strands that used to take part in R-loops were then ethanol-precipitated and reverse-transcribed into cDNA with the iScript Reverse Transcription Supermix (Bio-Rad) using a mix of poly(T) and random hexamers. Second-strand synthesis was performed using deoxyuridine triphosphate (dUTP) instead of deoxythymidine triphosphate (dTTP). After verifying the enrichments of R-loop (+) versus R-loop (−) loci by qPCR on the newly synthesized cDNA, sequencing libraries were built following sonication to reduce the size of cDNA fragments to ~200 bp. A uracil DNA glycosylase (UDG) step was added before the PCR amplification step to ensure strand specificity. Sequencing libraries were checked on an Agilent BioAnalyzer prior to sequencing on an Illumina HiSeq 2000. Because of size selection steps during library construction, the hybrids must be >50 bp in length to be detected. Read mapping after quality filters was performed using standard pipelines. Peak calling was performed using a custom-built hidden Markov model (Supplemental Experimental Procedures).

Measurement of R-loop Dynamics

Log-phase Ntera 2 cells were treated with 80 μM DRB, and samples were withdrawn at time zero (pre-treatment) and 5, 10, 20, 30, 45, 60, and 120 min after treatment and processed for DRIP followed by qPCR to interrogate specific promoter or terminal regions. All values were relative to input and normalized to time zero. After 120 min of DRB treatment, cells were washed twice in pre-warmed normal growth medium, and R-loop formation was again analyzed as a function of time.

Chromatin Signal Mapping, Enrichment, and Conservation

ENCODE and other datasets for histone modifications and chromatin binding factors are described in Table S1. Conservation of R-loop formation was analyzed over regions of human-mouse synteny identified by LiftOver. Detailed methods can be found in the Supplemental Experimental Procedures.

ACCESSION NUMBERS

The accession number for the DRIP-seq, DRIPc-seq, RNA-seq, and MethylC-seq datasets reported in this paper is NCBI GEO: GSE70189.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures, six figures, and two tables and can be found with this article online at <http://dx.doi.org/10.1016/j.molcel.2016.05.032>.

AUTHOR CONTRIBUTIONS

L.A.S., X.X., and P.A.G. performed the experimental work. S.R.H., Y.W.L., S.S., and A.R. performed the computational analysis. F.C. conceived the study, oversaw the work, and wrote the manuscript along with all other authors.

ACKNOWLEDGMENTS

We thank members of the F.C. lab for critical reading of the manuscript. This work was supported by NIH Grant GM094299 (to F.C.). L.A.S. was supported in part by a postdoctoral grant from the Philippe Foundation. S.R.H. is a Howard Hughes Medical Institute international student research fellow. Y.W.L. was supported in part by pre-doctoral training grant 5T32GM007377 and a dissertation year fellowship from University of California Davis. This work used the Vincent J. Coates Genomics Sequencing Laboratory at University of California Berkeley, supported by NIH S10 Instrumentation Grants S10RR029668 and S10RR027303.

Received: August 14, 2015

Revised: January 11, 2016

Accepted: May 23, 2016

Published: June 30, 2016

REFERENCES

- Aguilera, A., and García-Muse, T. (2012). R loops: from transcription byproducts to threats to genome stability. *Mol. Cell* 46, 115–124.
- Bell, A.C., West, A.G., and Felsenfeld, G. (1999). The protein CTCF is required for the enhancer blocking activity of vertebrate insulators. *Cell* 98, 387–396.
- Belotserkovskii, B.P., and Hanawalt, P.C. (2011). Anchoring nascent RNA to the DNA template could interfere with transcription. *Biophys. J.* 100, 675–684.
- Boguslawski, S.J., Smith, D.E., Michalak, M.A., Mickelson, K.E., Yehle, C.O., Patterson, W.L., and Carrico, R.J. (1986). Characterization of monoclonal antibody to DNA:RNA and its application to immunodetection of hybrids. *J. Immunol. Methods* 89, 123–130.
- Boque-Sastre, R., Soler, M., Oliveira-Mateos, C., Portela, A., Moutinho, C., Sayols, S., Villanueva, A., Esteller, M., and Guil, S. (2015). Head-to-head antisense transcription and R-loop formation promotes transcriptional activation. *Proc. Natl. Acad. Sci. USA* 112, 5785–5790.

- Carninci, P., Kasukawa, T., Katayama, S., Gough, J., Frith, M.C., Maeda, N., Oyama, R., Ravasi, T., Lenhard, B., Wells, C., et al.; FANTOM Consortium; RIKEN Genome Exploration Research Group and Genome Science Group (Genome Network Project Core Group) (2005). The transcriptional landscape of the mammalian genome. *Science* 309, 1559–1563.
- Castellano-Pozo, M., Santos-Pereira, J.M., Rondón, A.G., Barroso, S., Andújar, E., Pérez-Alegre, M., García-Muse, T., and Aguilera, A. (2013). R loops are linked to histone H3 S10 phosphorylation and chromatin condensation. *Mol. Cell* 52, 583–590.
- Chen, P.B., Chen, H.V., Acharya, D., Rando, O.J., and Fazio, T.G. (2015). R loops regulate promoter-proximal chromatin architecture and cellular differentiation. *Nat. Struct. Mol. Biol.* 22, 999–1007.
- Chu, Y., Simic, R., Warner, M.H., Arndt, K.M., and Prelich, G. (2007). Regulation of histone modification and cryptic transcription by the Bur1 and Paf1 complexes. *EMBO J.* 26, 4646–4656.
- de Jong, R.N., Truffault, V., Diercks, T., Ab, E., Daniels, M.A., Kaptein, R., and Folkers, G.E. (2008). Structure and DNA binding of the human Rtf1 Plus3 domain. *Structure* 16, 149–159.
- Dunn, K., and Griffith, J.D. (1980). The presence of RNA in a double helix inhibits its interaction with histone protein. *Nucleic Acids Res.* 8, 555–566.
- El Hage, A., Webb, S., Kerr, A., and Tollervey, D. (2014). Genome-wide distribution of RNA-DNA hybrids identifies RNase H targets in tRNA genes, retrotransposons and mitochondria. *PLoS Genet.* 10, e1004716.
- Ernst, J., and Kellis, M. (2012). ChromHMM: automating chromatin-state discovery and characterization. *Nat. Methods* 9, 215–216.
- Frietze, S., O'Geen, H., Blahnik, K.R., Jin, V.X., and Farnham, P.J. (2010). ZNF274 recruits the histone methyltransferase SETDB1 to the 3' ends of ZNF genes. *PLoS ONE* 5, e15082.
- Ginno, P.A., Lott, P.L., Christensen, H.C., Korf, I., and Chédin, F. (2012). R-loop formation is a distinctive characteristic of unmethylated human CpG island promoters. *Mol. Cell* 45, 814–825.
- Ginno, P.A., Lim, Y.W., Lott, P.L., Korf, I., and Chédin, F. (2013). GC skew at the 5' and 3' ends of human genes links R-loop formation to epigenetic regulation and transcription termination. *Genome Res.* 23, 1590–1600.
- Giresi, P.G., Kim, J., McDaniel, R.M., Iyer, V.R., and Lieb, J.D. (2007). FAIRE (Formaldehyde-Assisted Isolation of Regulatory Elements) isolates active regulatory elements from human chromatin. *Genome Res.* 17, 877–885.
- Gómez-González, B., García-Rubio, M., Bermejo, R., Gaillard, H., Shirahige, K., Marín, A., Foiani, M., and Aguilera, A. (2011). Genome-wide function of THO/TREX in active genes prevents R-loop-dependent replication obstacles. *EMBO J.* 30, 3106–3119.
- Grzechnik, P., Tan-Wong, S.M., and Proudfoot, N.J. (2014). Terminate and make a loop: regulation of transcriptional directionality. *Trends Biochem. Sci.* 39, 319–327.
- Hamperl, S., and Cimprich, K.A. (2014). The contribution of co-transcriptional RNA:DNA hybrid structures to DNA damage and genome instability. *DNA Repair (Amst.)* 19, 84–94.
- Heintzman, N.D., Stuart, R.K., Hon, G., Fu, Y., Ching, C.W., Hawkins, R.D., Barrera, L.O., Van Calcar, S., Qu, C., Ching, K.A., et al. (2007). Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nat. Genet.* 39, 311–318.
- Jonkers, I., and Lis, J.T. (2015). Getting up to speed with transcription elongation by RNA polymerase II. *Nat. Rev. Mol. Cell Biol.* 16, 167–177.
- Jonkers, I., Kwak, H., and Lis, J.T. (2014). Genome-wide dynamics of Pol II elongation and its interplay with promoter proximal pausing, chromatin, and exons. *eLife* 3, e02407.
- Klose, R.J., Yamane, K., Bae, Y., Zhang, D., Erdjument-Bromage, H., Tempst, P., Wong, J., and Zhang, Y. (2006). The transcriptional repressor JHDM3A demethylates trimethyl histone H3 lysine 9 and lysine 36. *Nature* 442, 312–316.
- Krajewski, W.A., Nakamura, T., Mazo, A., and Canaani, E. (2005). A motif within SET-domain proteins binds single-stranded nucleic acids and transcribed and supercoiled DNAs and can interfere with assembly of nucleosomes. *Mol. Cell Biol.* 25, 1891–1899.
- Kwak, H., Fuda, N.J., Core, L.J., and Lis, J.T. (2013). Precise maps of RNA polymerase reveal how promoters direct initiation and pausing. *Science* 339, 950–953.
- Lim, Y.W., Sanz, L.A., Xu, X., Hartono, S.R., and Chédin, F. (2015). Genome-wide DNA hypomethylation and RNA:DNA hybrid accumulation in Aicardi-Goutières syndrome. *eLife* 16, 4.
- Liu, W., Tanasa, B., Tyurina, O.V., Zhou, T.Y., Gassmann, R., Liu, W.T., Ohgi, K.A., Benner, C., Garcia-Bassets, I., Aggarwal, A.K., et al. (2010). PHF8 mediates histone H4 lysine 20 demethylation events involved in cell cycle progression. *Nature* 466, 508–512.
- Mayer, A., di Iulio, J., Maleri, S., Eser, U., Vierstra, J., Reynolds, A., Sandstrom, R., Stamatoyannopoulos, J.A., and Churchman, L.S. (2015). Native elongating transcript sequencing reveals human transcriptional activity at nucleotide resolution. *Cell* 161, 541–554.
- Noy, A., Pérez, A., Márquez, M., Luque, F.J., and Orozco, M. (2005). Structure, recognition properties, and flexibility of the DNA:RNA hybrid. *J. Am. Chem. Soc.* 127, 4910–4920.
- Parelho, V., Hadjur, S., Spivakov, M., Leleu, M., Sauer, S., Gregson, H.C., Jarmuz, A., Canzonetta, C., Webster, Z., Nesterova, T., et al. (2008). Cohesins functionally associate with CTCF on mammalian chromosome arms. *Cell* 132, 422–433.
- Powell, W.T., Coulson, R.L., Gonzales, M.L., Crary, F.K., Wong, S.S., Adams, S., Ach, R.A., Tsang, P., Yamada, N.A., Yasui, D.H., et al. (2013). R-loop formation at Snord116 mediates topotecan inhibition of Ube3a-antisense and allele-specific chromatin decondensation. *Proc. Natl. Acad. Sci. USA* 110, 13938–13943.
- Ratmeyer, L., Vinayak, R., Zhong, Y.Y., Zon, G., and Wilson, W.D. (1994). Sequence specific thermodynamic and structural properties for DNA:RNA duplexes. *Biochemistry* 33, 5298–5304.
- Shilatfard, A. (2012). The COMPASS family of histone H3K4 methylases: mechanisms of regulation in development and disease pathogenesis. *Annu. Rev. Biochem.* 81, 65–95.
- Skourti-Stathaki, K., and Proudfoot, N.J. (2014). A double-edged sword: R loops as threats to genome integrity and powerful regulators of gene expression. *Genes Dev.* 28, 1384–1396.
- Skourti-Stathaki, K., Proudfoot, N.J., and Gromak, N. (2011). Human senataxin resolves RNA/DNA hybrids formed at transcriptional pause sites to promote Xrn2-dependent termination. *Mol. Cell* 42, 794–805.
- Skourti-Stathaki, K., Kamieniarz-Gdula, K., and Proudfoot, N.J. (2014). R-loops induce repressive chromatin marks over mammalian gene terminators. *Nature* 516, 436–439.
- Stirling, P.C., Chan, Y.A., Minaker, S.W., Aristizabal, M.J., Barrett, I., Sipahimalani, P., Kobor, M.S., and Hieter, P. (2012). R-loop-mediated genome instability in mRNA cleavage and polyadenylation mutants. *Genes Dev.* 26, 163–175.
- Tomson, B.N., and Arndt, K.M. (2013). The many roles of the conserved eukaryotic Paf1 complex in regulating transcription, histone modifications, and disease states. *Biochim. Biophys. Acta* 1829, 116–126.
- Wahba, L., Amon, J.D., Koshland, D., and Vuica-Ross, M. (2011). RNase H and multiple RNA biogenesis factors cooperate to prevent RNA:DNA hybrids from generating genome instability. *Mol. Cell* 44, 978–988.
- Yu, K., Chedin, F., Hsieh, C.L., Wilson, T.E., and Lieber, M.R. (2003). R-loops at immunoglobulin class switch regions in the chromosomes of stimulated B cells. *Nat. Immunol.* 4, 442–451.